

Research Article

Benchmarking three DNA metabarcoding technologies for efficient detection of non-native cerambycid beetles in trapping collections

Loïs Veillat¹, Stéphane Boyer², Marina Querejeta^{2,3}, Emmanuelle Magnoux¹, Alain Roques¹, Carlos Lopez-Vaamonde^{1,2*}, Geraldine Roux^{1,4*}

¹ INRAE, UR633, Zoologie Forestière, F-45075 Orléans, France

² Institut de Recherche sur la Biologie de l'Insecte, UMR7261 CNRS - Université de Tours, Avenue Monge, 37200 Tours, France

³ UMR CNRS 7267, Ecologie et Biologie des Interactions, Université de Poitiers, 3 Rue Jacques Fort, 86073, Poitiers, France

⁴ Laboratoire Physiologie, Ecologie et Environnement P2E, Université d'Orléans, 1 rue de Chartres, B.P. 6759, 45067 Orléans cedex 2, France

Corresponding authors: Loïs Veillat (lois.veillat@outlook.com); Geraldine Roux (geraldine.roux@univ-orleans.fr)

Abstract

Individual sorting and identification of thousands of insects collected in mass trapping biosurveillance programmes is a labour-intensive and time-consuming process. Metabarcoding allows the simultaneous identification of multiple individuals in a single mixed sample and has the potential to expedite this process. However, detecting all the species present in a bulk sample can be challenging, especially when under-represented non-native specimens are intercepted.

In this study, we quantified the effectiveness of DNA metabarcoding at detecting exotic species within six different mock communities of native and non-native species of European xylophagous cerambycid beetles. The main objective is to compare three different sequencing technologies (MinION, Illumina and IonTorrent) to evaluate which one is the most suitable in this context. Additionally, dry and wet (monopropylene glycol and water) collection methods were compared. Although not observing significant differences in the total number of species detected amongst the three sequencing technologies, the MinION detected a greater number of species in field-like samples. All three sequencing technologies achieved success in detecting and identifying closely-related species and species in low abundance. The capture method of insects in the field greatly influenced sample preservation and detection. Individuals captured in traps containing monopropylene and water had lower DNA concentration, leading to lower species detection rates compared to individuals killed using just an insecticide without any collection medium.

Key words: Alien, biological invasions, biosecurity, Cerambycidae, exotic, Illumina®, IonTorrent®, Oxford Nanopore®, xylophagous



Academic editor: Marcela Uliano-Silva

Received: 21 June 2024

Accepted: 22 October 2024

Published: 25 November 2024

Citation: Veillat L, Boyer S, Querejeta M, Magnoux E, Roques A, Lopez-Vaamonde C, Roux G (2024) Benchmarking three DNA metabarcoding technologies for efficient detection of non-native cerambycid beetles in trapping collections. NeoBiota 96: 237–259. <https://doi.org/10.3897/neobiota.96.130195>

Copyright: © Loïs Veillat et al.

This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

Introduction

The exponential increase in biological invasions that has been observed over the past decades is expected to persist (Seebens et al. 2021). This is primarily due to factors such as globalisation, tourism and global warming (Chown et al. 2015). Amongst the species introduced beyond their native range by human activities, insects are the most prevalent group (Seebens et al. 2018) and can cause a wide range of impacts. Non-native insects can affect native flora, fauna and ecosystems in various ways

* These authors contributed equally to this work.

(Kenis et al. 2009) and they can also transmit pathogens and diseases, thus threatening public health (Mazza et al. 2014). Economic implications are also to be considered since numerous invasive insects are important pests for agricultural crops and plantation forests, inducing huge management costs (Bradshaw et al. 2016).

Amongst these non-native insects, species associated with woody plants are increasingly dominating, accounting for 76.5% of all herbivore species newly recorded in Europe from 2000 to 2014, potentially because of the growing trade of ornamental plants and wooden packaging material transported in international cargo shipments (Aukema et al. 2010; Roques et al. 2016). One of these important families of xylophagous beetles is the long-horned Cerambycidae, with more than 200 species affecting forestry, horticulture and agriculture (Rossa and Goczał 2021), resulting in multimillion-dollar losses every year (Wang 2017). To detect potential new invasions of Cerambycids, biomonitoring programmes have been set up over large geographical areas with intensive trapping campaigns extending over several years (Roques et al. 2023; Mas et al. 2023). However, rapidly evolving trades lead to changes in trade routes and imported goods, which results in an increasing arrival of new non-native species. A high number of these species have not been previously reported as invaders; some are not considered to be pests in their native ranges and some could even be unknown to science (Seebens et al. 2018). As part of the European project HOMED (<https://homed-project.eu/>), 244 Cerambycid traps were set up across Europe (France, Italy, Spain, Switzerland, Portugal, Austria, England, Greece, Slovenia, Netherlands, Bulgaria, Czech Republic and Sweden), 38 in Asia (China, Siberia, Russia), 11 in North America (USA, Canada), five in the Caribbean (Martinique) and four in Australia, all baited with generic lures for simultaneous detection of multiple species (Roques et al. 2023). In such large-scale trapping campaigns, thousands of captured insects must be sorted and identified by expert taxonomists. This identification step is time-consuming and labour-intensive, thus limiting the rapid detection of non-native individuals amongst large numbers of native ones (Piper et al. 2019; Abeynayake et al. 2021; Chua et al. 2023). Yet, it is essential that those non-native species are detected as quickly as possible to allow their eradication before establishment and dispersal (Richardson et al. 2000; Blackburn et al. 2011; Giovani et al. 2020).

For insects, traditional DNA barcoding, using a short fragment of the Cytochrome Oxidase 1 (*COI*) gene, has truly become a universal tool to identify unknown specimens at species level regardless of sex or life stage (Hebert et al. 2003). Namely, DNA barcoding has been successfully used to accurately identify cerambycid pest species for biomonitoring (Hodgetts et al. 2016; Wu et al. 2017; Kelnarova et al. 2019; Javal et al. 2021). Despite its numerous advantages, individual DNA barcoding remains a laborious and time-consuming approach in the context of mass-trapped insects, as it requires individual sorting of thousands of specimens, tissue sampling (often legs), DNA extraction and amplification and finally sequencing of each sample individually. However, the recent application of high-throughput sequencing (HTS) technologies to DNA barcoding allows for the expedited production of thousands of DNA barcodes (deWaard et al. 2019; Srivathsan et al. 2021).

This metabarcoding approach generates a large number of short DNA sequences (reads), allowing the accurate identification of multiple species simultaneously from a single mixed sample (hereafter called “bulk”) (Liu et al. 2020), such as all the individual insects captured in a single biomonitoring trap. Moreover, compared to traditional morphological identification, metabarcoding offers a significant reduction in costs (Batovska et al. 2021), generally providing equivalent or better

detection and identifying a much wider spectrum of taxa (Elbrecht et al. 2017; Andújar et al. 2018). Using DNA as a proxy for species detection and considering sequence variation within and amongst taxa, metabarcoding approaches are, however, constrained by the completeness of the reference databases to accurately assign sequences to correctly identified taxa (Liu et al. 2020).

Although metabarcoding has several advantages, ensuring the accuracy of detections is crucial. Erroneous detections of pest species can have severe environmental and economic consequences (Batovska et al. 2021). Yet, metabarcoding approaches still suffer from methodological limitations that may make them unsuitable for rapid biosecurity detection (contaminations, unreliable quantification, incomplete databases, false positives etc.). One specific challenge is the time required to process samples, which can suffer long delays between capturing individuals and obtaining sequencing results. This is especially true when sampling sites are located far away from laboratories or when transporting samples requires specific permits for certain species. Additionally, when external providers are slow to sequence samples, the process is further delayed. These limitations can hinder biomonitoring projects and slow down the detection of potential invasive species. As a result, the implementation of measures to mitigate their impacts may also be delayed (Krehenwinkel et al. 2019; Egeter et al. 2022). Despite these limitations, the Illumina MiSeq sequencing technology has been favoured due to its lower error rate and well-established bioinformatic procedures (Piper et al. 2019). Yet, Braukmann et al. (2019) demonstrated similar performance in sequence quality and insect species recovery using IonTorrent platforms (Ion Torrent PGM and Ion Torrent S5). As these machines are more affordable than the MiSeq, they can easily be purchased by individual laboratories. Hence, this technology is overall less often dependent on external providers and their sequencing delays.

Over the past few years, Oxford Nanopore Technologies® has released a very inexpensive portable sequencing platform, the MinION. This small sequencer can be connected via USB to a laptop to perform sequencing (Krehenwinkel et al. 2019) in the field and obtain sequencing data in real time conditions. Indeed, the MinION for a metabarcoding application offers the possibility of performing DNA sequencing of bulk samples directly on site without the need for transport or reliance on external sequencing providers. So far, although the MinION does not seem suitable for the characterisation of complex communities, it is already suitable for the analysis of metabarcoding data when the species diversity per sample is low and the target species are well represented in public databases (Ho et al. 2020). In addition, recent developments in Nanopore technology and base calling have reduced sequence error from 6% (Srivathsan et al. 2021) down to less than 1% (Srivathsan et al. 2024).

The primary objective of our study was to determine the most effective metabarcoding approach for the biosurveillance of Cerambycid wood-boring beetles. To achieve this, we compared the performance of three Next Generation Sequencing technologies: the portable Nanopore sequencer MinION, the Illumina MiSeq and the Ion GeneStudio S5 (IonTorrent®). Our evaluation focused on their ability to detect invasive species in different mock communities. Specifically, we assessed their accuracy in differentiating between closely-related cerambycid species and detecting low-abundance species in mixed-trap samples. Additionally, we analysed various metabarcoding primer pairs to evaluate their accuracy in species identification. Finally, we emphasised the significance of the field sampling protocol, particularly the trapping methods (dry versus monopropylene glycol) in species detection.

Materials and methods

Taxa sampling

Mock communities were constructed using 48 field-trapped specimens from different countries in Europe (France, Greece, Portugal, Spain), China (Beijing and Zhejiang Province) and the USA (Michigan) (Table 1), as part of a worldwide trapping experiment using multi-funnel traps baited with a generic attractant blend, including eight Cerambycid pheromones (see details of the blend composition and trapping methods in Roques et al. (2023)). Most of the specimens (36/48) were caught using α -cypermethrin insecticide (Storanet®, BASF Pflanzenschutz Deutschland, Germany) in the trap basins, of which the bottom had been replaced with a wire mesh to allow drainage and keep specimens dry (hereafter called the “dry” method). Other specimens (12/48) were captured using a 50:50 ratio of monopropylene glycol (MPG) and water (hereafter called the “wet” method). Cerambycides collected from field-traps were stored in ethanol 95% and kept at -20 °C until further processing. Two individuals were captured by hand (“hand collected” in Table 1) and pinned in collection boxes after capture. The date, country of collection, type of trap and the 48 specimens used in mock communities are detailed in Table 1.

Mock community construction and DNA extraction

Six mock communities with varying species composition were assembled as follows:

Test 1: Identifying closely-related species.

To assess the efficiency of the different sequencing technologies and primers to differentiate between sister species, bulks 1 and 2 were composed of congeneric species (Table 1). Two legs from each individual (one specimen per species) were collected and pooled to constitute the bulks. The whole set of legs was then ground using flame-sterilised metal pestles to limit the risk of contamination. DNA from the ground material was extracted using the DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer’s instructions. Two additional legs were taken from the same specimens to assess the quantity and quality (A260/280 and A260/230 ratios) of DNA for each specimen individually (Fig. 1a).

Test 2: Detecting low abundance species.

Bulks 3 and 4 were composed of six species represented by heterogeneous DNA concentrations (Suppl. material 1: table S1) to assess the ability of the sequencing technologies and primers to detect species present in very low abundance. DNA of each individual (one specimen per species) was previously extracted using two legs that were ground as above and processed using the Qiagen DNeasy Blood and Tissue Kit. To construct bulks 3 and 4, individual DNA extracts were quantified using a fluorometer (Nanodrop™, Thermo Fisher Scientific) and mixed together according to their concentration to achieve the needed proportions of DNA for each individual (six individuals of different species ranging from 41% to 3% for Bulk 3 and six individuals of different species ranging from 50% to 0.5% for Bulk 4) (Table 1, Fig. 1b).

Table 1. Species, origin, date and condition of capture of the specimens used in the six bulks. Species names in bold correspond to exotic species. We consider specimens that have been captured on a different continent from their place of origin as exotic.

Bulk	Species	Country of collection	Collection Year	Collection type
1	<i>Arhopalus ferus</i>	Portugal	2020	Cypermethrin insecticide (dry method)
1	<i>Arhopalus rusticus</i>	France	2021	Cypermethrin insecticide (dry method)
1	<i>Arhopalus syriacus</i>	Portugal	2019	Monopropylene glycol (wet method)
1	<i>Xylotrechus arvicola</i>	Portugal	2021	Cypermethrin insecticide (dry method)
1	<i>Xylotrechus chinensis</i>	Greece	2019	Cypermethrin insecticide (dry method)
1	<i>Xylotrechus stebbingi</i>	Greece	2019	Cypermethrin insecticide (dry method)
1	<i>Xylotrechus undulatus</i>	USA	2019	Monopropylene glycol (wet method)
2	<i>Monochamus galloprovincialis</i>	Portugal	2019	Monopropylene glycol (wet method)
2	<i>Monochamus sutor</i>	France	2019	Cypermethrin insecticide (dry method)
2	<i>Monochamus carolinensis</i>	USA	2019	Monopropylene glycol (wet method)
2	<i>Monochamus scutellatus</i>	USA	2019	Monopropylene glycol (wet method)
2	<i>Phymatodes amoenus</i>	USA	2019	Monopropylene glycol (wet method)
2	<i>Phymatodes testaceus</i>	USA	2019	Monopropylene glycol (wet method)
2	<i>Phymatodes varius</i>	USA	2019	Monopropylene glycol (wet method)
2	<i>Phymatodes aereus</i>	USA	2019	Monopropylene glycol (wet method)
2	<i>Phymatodes dimidiatus</i>	USA	2019	Monopropylene glycol (wet method)
3	<i>Pyrrhidium sanguineum</i>	France	2020	Cypermethrin insecticide (dry method)
3	<i>Xylotrechus stebbingi</i>	Spain	2021	Cypermethrin insecticide (dry method)
3	<i>Monochamus galloprovincialis</i>	Spain	2021	Cypermethrin insecticide (dry method)
3	<i>Xylotrechus chinensis</i>	Greece	2019	Cypermethrin insecticide (dry method)
3	<i>Chlorophorus glabromaculatus</i>	France	2020	Cypermethrin insecticide (dry method)
3	<i>Phymatodes testaceus</i>	France	2020	Cypermethrin insecticide (dry method)
4	<i>Arhopalus ferus</i>	France	2020	Cypermethrin insecticide (dry method)
4	<i>Monochamus sutor</i>	France	2019	Cypermethrin insecticide (dry method)
4	<i>Aegomorphus francottei</i>	France	2020	Cypermethrin insecticide (dry method)
4	<i>Monochamus galloprovincialis</i>	France	2018	Cypermethrin insecticide (dry method)
4	<i>Xylotrechus stebbingi</i>	Spain	2021	Cypermethrin insecticide (dry method)
4	<i>Xylotrechus chinensis</i>	Greece	2019	Cypermethrin insecticide (dry method)
5	<i>Pyrrhidium sanguineum</i>	France	2021	Cypermethrin insecticide (dry method)
5	<i>Batocera rubus</i>	China	2012	Hand collected
5	<i>Cerambyx scopolii</i>	France	2020	Cypermethrin insecticide (dry method)
5	<i>Cordylomera spinicornis</i>	France	2020	Cypermethrin insecticide (dry method)
5	<i>Leiopus femoratus</i>	France	2021	Cypermethrin insecticide (dry method)
5	<i>Leiopus nebulosus</i>	France	2020	Cypermethrin insecticide (dry method)
5	<i>Pachyta bicuneata</i>	China	1987	Hand collected
5	<i>Stictoleptura cordigera</i>	France	2021	Cypermethrin insecticide (dry method)
6	<i>Arhopalus rusticus</i>	France	2020	Cypermethrin insecticide (dry method)
6	<i>Xylotrechus chinensis</i>	Greece	2019	Cypermethrin insecticide (dry method)
6	<i>Plagionotus detritus</i>	France	2020	Cypermethrin insecticide (dry method)
6	<i>Plagionotus arcuatus</i>	France	2020	Cypermethrin insecticide (dry method)
6	<i>Xylotrechus stebbingi</i>	France	2020	Cypermethrin insecticide (dry method)
6	<i>Arhopalus syriacus</i>	France	2020	Cypermethrin insecticide (dry method)
6	<i>Arhopalus ferus</i>	France	2020	Cypermethrin insecticide (dry method)
6	<i>Xylotrechus colonus</i>	USA	2019	Monopropylene glycol (wet method)
6	<i>Chlorophorus ruficornis</i>	France	2021	Cypermethrin insecticide (dry method)
6	<i>Phymatodes testaceus</i>	France	2021	Cypermethrin insecticide (dry method)
6	<i>Prionus coriarius</i>	France	2010	Cypermethrin insecticide (dry method)
6	<i>Phymatodes amoenus</i>	USA	2019	Monopropylene glycol (wet method)

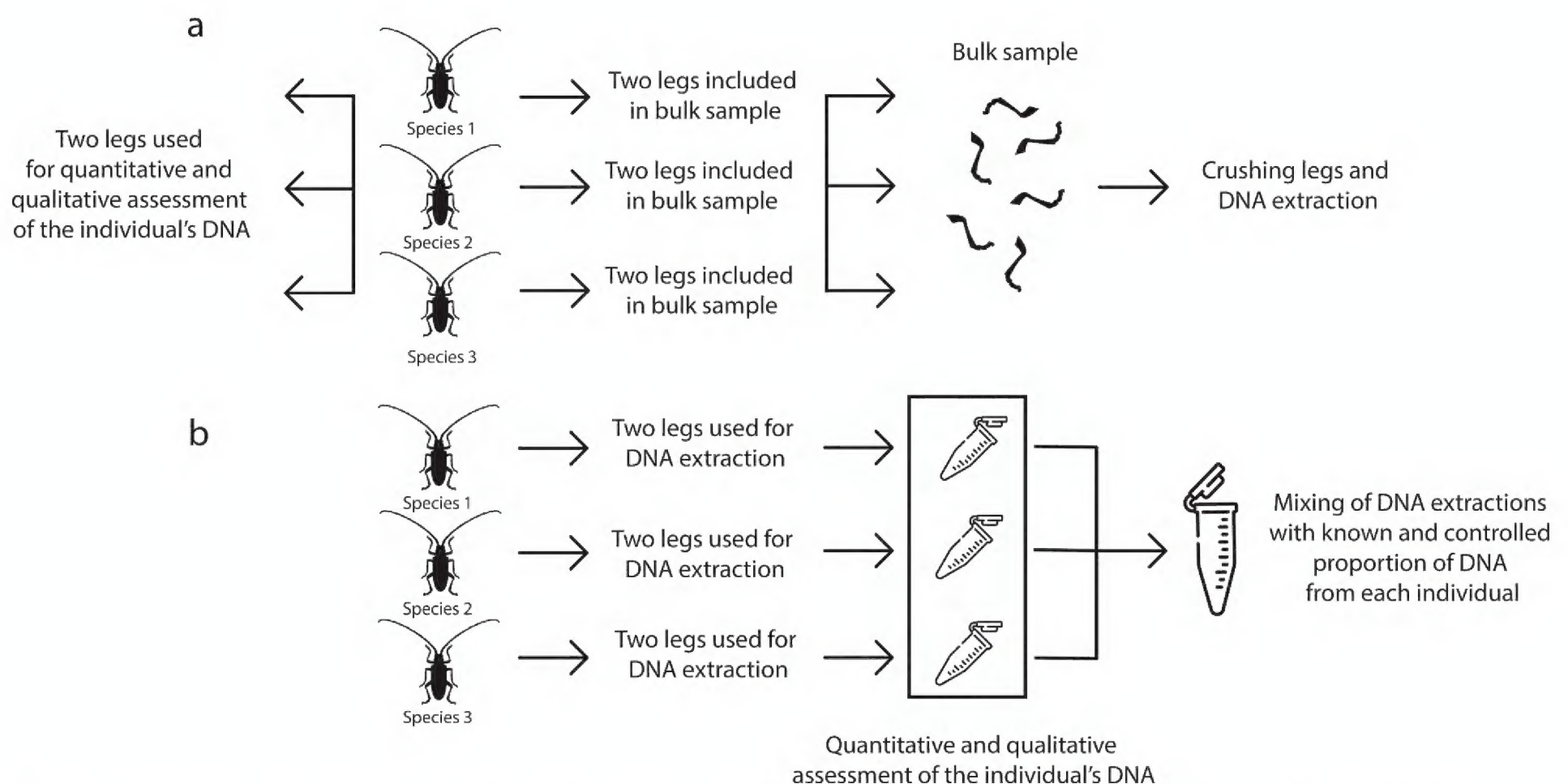


Figure 1. Overview of the DNA extraction protocol for tests 1 (identifying closely-related species) and 3 (mimicking field trap content on species composition) (a) and for test 2 (detecting low abundance species) (b).

Test 3: Mimicking field trap content on species composition.

Bulks 5 and 6 were built to reconstitute real trap contents by a collaborator involved in Cerambycidae trapping campaigns using multi-pheromonal traps (Roques et al. 2023). These bulks include individuals from a number of species native to Europe usually found in the traps deployed there, with the addition of non-native species that have either already been introduced or are at risk of being introduced in Europe (Bulk 5: 22 individuals of eight species, including one non-native (*Cordylomera spinicornis*); Bulk 6: 41 individuals of 12 species including two non-native ones (*Xylotrechus stebbingi* and *Xylotrechus chinensis*) (Table 1). The DNA was extracted following the same protocol as for bulks 1 and 2 where two legs were taken from each individual and ground together for DNA extraction (Fig. 1a).

PCR amplification

All bulk samples were amplified with two pairs of primers internal to the commonly used barcode fragment: BF3/BR2 (called hereafter “B”) (CCHGAYATRGC-HTTYCCHCG / TCDGGRTGNCCRAARAAYCA (Elbrecht and Leese 2017; Elbrecht et al. 2019), which generates a 458 bp amplicon that was used for all the technologies; and fwhF2/fwhR2n (called hereafter “F”) (GGDACWGGWT-GAACWGTWTAYCCHCC / GTRATWGCHCCDGCTARWACWGG), which generates a shorter 254 bp amplicon (Vamos et al. 2017) and was used for Illumina and MinION technologies only. Each PCR comprised 15.3 µl H₂O, 2.5 µl 10X PCR buffer, 2.5 µl dNTPs [1 mM], 1 µl of each primer [0.4 mM], 0.2 µl Dream Taq (Thermo Fisher Scientific), 0.5 µl Betaine [100 mM] and 2 µl DNA for a total of 25 µl per reaction. For both primer pairs, PCR was performed using the same programme: 95 °C for 5 min, 29 cycles of 95 °C for 30 s, 48 °C for 30 s and 72 °C for 50 s, followed by 72 °C for 5 min (Elbrecht et al. 2019). PCR products were then run on a 2% agarose gel stained with ethidium bromide and

visualised by UV transilluminator. The PCR products were then purified with the NucleoFast 96 PCR plate purification kit (Macherey-Nagel). Three PCR replicates were performed for the six bulks when using the MinION technology.

Illumina® library preparation

A second ligation PCR was performed on the products of the first PCR to add Illumina® tags and adapters, prepared by ligating Nextera XT indices through an eight cycle PCR (with a modified PCR protocol). The second PCR was carried out with the same conditions as for the initial PCR. Reactions (25 µl) contained the following: 5 µl of template DNA (purified products from the first PCR), 1 µl of each primer [10 µM], 5 µl of 5X GoTaq (Promega) reaction buffer, 1 µl of MgCl₂ [25 mM], 1 µl of BSA [1 mg/ml], 0.5 µl of dNTPs [5 mM], 0.125 µl of GoTaq G2 Polymerase (Promega) and 10.375 µl of molecular-grade water to reach 25 µl. The PCR conditions were the same as for the first PCR, with eight cycles. The products of the second PCR were verified on a 2% agarose gel. PCR products were then equimolarly pooled into two different pools (one pool per primer pair used) and purified using the GeneJET Gel Extraction kit from an agarose gel, following the manufacturer's instructions. This library was sequenced in Illumina MiSeq using V3 chemistry (300 × 300 bp, 600 cycles) in the Sequencing Center within the Biozentrum of the Ludwig-Maximilian University in Munich (Germany).

MinION library preparation

Libraries were prepared according to the Oxford Nanopore Technologies® protocol: “PCR barcoding (96) amplicons (SQK-LSK110) (version: PBAC96_9114_v110_revF_10Nov2020)” with the following specifications. After the first PCR described above, the Nanopore PCR barcoding expansion Pack 1-96 (EXP-PBC096) was used to perform the second PCR to incorporate the Oxford Nanopore Technologies® barcode sequences on the amplicons generated in the first PCR.

Reactions (50 µl) contained the following: 2 µl of template DNA (purified products from the first PCR), 0.5 µl of each primer [10 µM], 10 µl of 5X GoTaq (Promega) reaction buffer, 2 µl of MgCl₂ [25 mM], 2 µl of BSA [1 mg/ml], 2 µl of Q solution, 1 µl of dNTPs [5 mM], 0.3 µl of GoTaq G2 Polymerase (Promega) and 29.7 µl of molecular-grade water to reach 25 µl. The thermocycling conditions followed the manufacturer recommendations: 95 °C for 3 min, followed by 15 cycles of 95 °C for 15 s, 62 °C for 15 s and 65 °C for 30 s and 65 °C for 7 min.

Final PCR products were then quantified using Qubit and equimolarly pooled before being purified with Agencourt AMPure XP beads (Beckam Coutler). The final pool was then sequenced on the MinION sequencer (Mk1c; Oxford Nanopore Technologies®, UK) using a R10.3 flowcell (MIN111) with 1331 pores available and the LSK110 ligation sequencing kit. The two replicates of bulk 6 using the MinION technology were of insufficient quality (Nanodrop) and were, therefore, removed from the analysis.

IonTorrent® library preparation

For the production of the libraries, we started with 5 ng of DNA extract (Qubit measurement). The Nextflex Cellfree DNaseq kit (PerkinElmer) was used for the process. The quality of the libraries was assessed using Qubit (for quantification)

and Bioanalyzer (using the HighSensitivity kit from Agilent, for size verification). After quality control, each library was amplified by emulsion PCR on the Ion One Touch 2 instrument, with a concentration of 15 pg/μl. Subsequently, the libraries were sequenced on an Ion GeneStudio S5 system using a single-end sequencing protocol with a 300 bp read length. Sequencing was performed on an Ion 520 Chip by the GeT-BioPuces platform (Toulouse, France).

False positive detections

False positive detections are considered to be the detection of a species within a bulk that was not initially present when the bulks were constructed. In order to estimate the representativeness of false positives within true positives in the bulks, the total number of reads assigned to false positive OTUs was reconciled and compared to the number of reads assigned to non-false positive detections. The number of false positives detected according to the different tested combinations is indicated in Suppl. material 1: table S2.

Reference barcode dataset

A dataset was built using all the public sequences of all Cerambycidae species available in BOLD systems v.4 (Ratnasingham and Hebert 2007). It was then verified whether all 33 species present in the bulk samples were represented by at least one sequence in the database. Three species not previously included in the database were barcoded through Sanger sequencing on an ABI 3500 genetic analyser (Applied Biosystems) using the Big-Dye Terminator V3.1 sequencing kit (Applied Biosystems) and BF3/BR2 primer pair. They were subsequently added to our local database to ensure that they were represented by at least one barcode sequence. The number of sequences in the database for each species is shown in Suppl. material 1: table S1.

The final reference dataset is available from BOLD in the dataset DS-MINION (dx.doi.org/10.5883/DS-MINION) and includes one barcode per species together with the three newly-generated barcodes. Lab Sheet from the DS-MinION database is shown in Suppl. material 1: table S3.

Illumina® data processing

The raw data were analysed using the FROGS v.4.0.1 pipeline, a standardised pipeline containing a set of tools that are used to process amplicon reads that have been produced from Illumina® sequencing (Escudié et al. 2018; Henrie et al. 2022). First, amplicons with a size between 408 and 508 for the BF3/BR2 primer pair and 204 and 304 for the fwhF2/fwhR2n primer pair were retained using the Pre-process tool. For this first step, paired-end reads are merged using VSEARCH (Rognes et al. 2016) with a mismatch rate set at 10%. Cutadapt (Martin 2011) is then used to remove sequences in which both primers are absent and to trim the primers. Sequence clustering was then performed using the SWARM algorithm (Mahé et al. 2014) with a maximum sequence difference set at $d = 1$ (--distance 1 parameter), as recommended by SWARM. Chimeric sequences were then removed with the Remove chimera tool relying on VSEARCH with the de novo UCHIME method (Edgar et al. 2011; Rognes et al. 2016). Sequences were aligned to the same

database used for the MinION and IonTorrent® data analysis. In order to remove all spurious detections, OTU detections with fewer than 10 reads were removed. In barcoding and metabarcoding studies of insects, the sequence similarity level for OTU identification usually ranges from 95% to 99% (e.g. Gibson et al. (2014); Zenker et al. (2016)). We calculated the best threshold value for our dataset by applying the function `localMinima` from the R package `spider` v.1.5.0 (Brown et al. 2012). Based on this analysis, we used a threshold of 98% to assign OTUs to species level with `blastn+` (Camacho et al. 2009). The resulting OTU tables for Illumina F and Illumina B are provided in Suppl. material 1: tables S4, S5, respectively.

MinION and IonTorrent® data processing

Bioinformatics analyses were performed on the Genotoul Bioinformatics Platform (INRAE, Toulouse, France). Basecalling and demultiplexing were performed for MinION data using Guppy v.6.1.7; ONT; high accuracy base calling mode; parameters: `-c dna_r10.3_450bps_hac.cfg --min_qscore 5 --trim_barcodes`. Then, for MinION and IonTorrent® data, we used the *msi* data processing pipeline v.0.3.6 (Egenter et al. 2022) to reduce the error rate of the reads by polishing them after the basecalling step. Reads smaller than 40 bp were removed with `cutadapt` v.4.0 (Martin 2011). The size range was set between 408 bp and 508 bp for BF3/BR2 and between 204 bp and 304 bp for fwhF2/fwhR2n. The clustering step was carried out with `ISONCLUST` v.0.0.6.1 (Sahlin and Medvedev 2020; with parameters: `--mapped_threshold 0.825` and `--aligned_threshold 0.55`) and a consensus sequence per cluster was generated using `RACON` v.1.5.0 (Vaser et al. 2017). The polished reads were then clustered at 97% sequence identity with `CD-HIT` v.4.8.1 (Fu et al. 2012) and a representative sequence from each cluster (centroid) was selected. The polished reads were then aligned to the local database with `BLAST` (BLASTn algorithm). The following parameters were used: `-word_size 11 -perc_identity 95 -qcov_hsp_perc 98 -gapopen 0 -gapextend 2 -reward 1 -penalty 1 -max_target_seqs 100`. Similarly, to the Illumina® data processing, OTU detections with less than 10 reads were removed. Finally, a taxonomic assignment was performed for each query using a Lowest Common Ancestor (LCA) approach with the bioinformatics package `metabinkit` (Chain et al. 2016; Egenter et al. 2018; Kitson et al. 2019) with the following thresholds: 98% at species level, 97% at genus level, 95% at family level (Alberdi et al. 2018; Egenter et al. 2022). The resulting OTU tables for MinION B, MinION F and IonTorrent B are provided in Suppl. material 1: tables S6–S8, respectively.

Statistical analysis

A two-sample test of proportions was used to compare and assess the significance of the proportion of reads assigned to the species levels for MinION, Illumina and IonTorrent technologies using the “Social Science Statistics” website (<https://www.socscistatistics.com/tests/anova/default2.aspx>). The proportion of reads assigned to different taxonomic levels was calculated by summing the total reads from different bulk samples for each condition. To determine if the number of false positives was significantly different amongst the three technologies and the two primer pairs, we calculated the detection mean for each bulk under different conditions. We then performed an ANOVA test followed by a Tukey HSD test using the “Social Science Statistics” website. The Wilcoxon test, Exact Fisher’s test and standard deviation

were calculated in R v.4.3.2 (<https://www.R-project.org/>). Sensibility, which measures the ability to correctly identify true positives, was calculated using the following formula: true positives / (true positives + false negatives) and precision, which measures the ability to measure the proportion of correct detections, was calculated using the following formula: true positives / (true positive + false positives).

Results

A total of 1,248,595 reads were sequenced with the MinION Nanopore® technology using the F primer pair, with an average of 78,037 (SD = 28,415) reads per sample. After quality filtering and removal of reads of incorrect size or insufficient quality, 1,113,844 (89.2%) reads were retained, with an average of 69,615 (SD = 25,508) reads per bulk. For the B primer pair, a total of 1,132,604 reads were sequenced, with an average of 62,922 (SD = 17,442) reads per sample. After quality filtering, a total of 948,832 (83.8%) reads were retained, with an average of 52,712 (SD = 14,512) reads per bulk (Table 2).

Table 2. Number of raw reads obtained after sequencing and after pre-processing steps according to sequencing technologies and primer pairs used.

Technology	Primer pair	n_raw_reads	n_reads_post_filtering
MinION	B	1,132,604	948,832
	F	1,248,595	1,113,844
Illumina	B	1,549,894	1,025,637
	F	2,383,028	1,686,058
IonTorrent	B	838,489	280,695

The Illumina® sequencing produced a total of 1,549,894 reads using the B primer pair, with an average of 258,316 (SD = 39,365) reads per bulk. After quality filtering, 1,025,637 (66.2%) reads were retained, with an average of 170,940 (SD = 69,961) reads per bulk. For the F primer pair, a total of 2,383,028 reads were sequenced, with an average of 397,171 (SD = 84,482) reads per bulk. After quality filtering, 1,686,058 (73.3%) reads were retained, with an average of 281,010 (SD = 112,512) reads per bulk (Table 2).

Regarding the IonTorrent® technology, 838,489 reads were sequenced, with an average of 139,748 (SD = 17,086) reads per bulk with the B primer pair. After quality filtering, 280,695 (33.5%) reads remain, with an average of 46,782 (SD = 5,025) reads per bulk (Table 2).

Benchmarking of sequencing technologies

The MinION technology accurately identified 28 out of 48 specimens at the species level, Illumina® technology allowed specific identification of 27 specimens and IonTorrent® identified 24 specimens. The primer pair F enabled the specific identification of 27 specimens at species level, while the primer pair B enabled the identification of 31 specimens at species level. Illumina® F, Illumina® B and MinION B allowed for 25 species-level identifications across all bulks (sensibility = 0.52) and 24 for MinION F and for IonTorrent® B (sensibility = 0.50). This difference was not significant (Fisher’s Exact Test, p = 1.00) (Fig. 2).

The number of reads assigned at the species level was significantly higher with Illumina® technology (p.value < 0.00001) compared to MinION. Nearly 97% of reads were assigned at the species level for the Illumina® F combination compared to 90% for the MinION F combination (p.value < 0.0001). As for primer pair BF3/BR2, over 87.3% of reads were assigned at the species level for Illumina®, followed by over 79.7% for MinION technology and over 77.2% for IonTorrent® technology (Fig. 3). The primer pair fwhF2/fwhR2n resulted in a significantly higher percentage of reads assigned at the species level (93.6%) (considering both Illumina® and MinION technologies) compared to the couple of primers B (81.4%) (considering all three technologies) (p. value < 0.00001). The summary of the number of reads is shown in Suppl. material 1: table S9a, while the percentage is shown in Suppl. material 1: table S9b. Both tables display data assigned to the taxonomic levels of species, genus, family or higher, for all conditions tested.

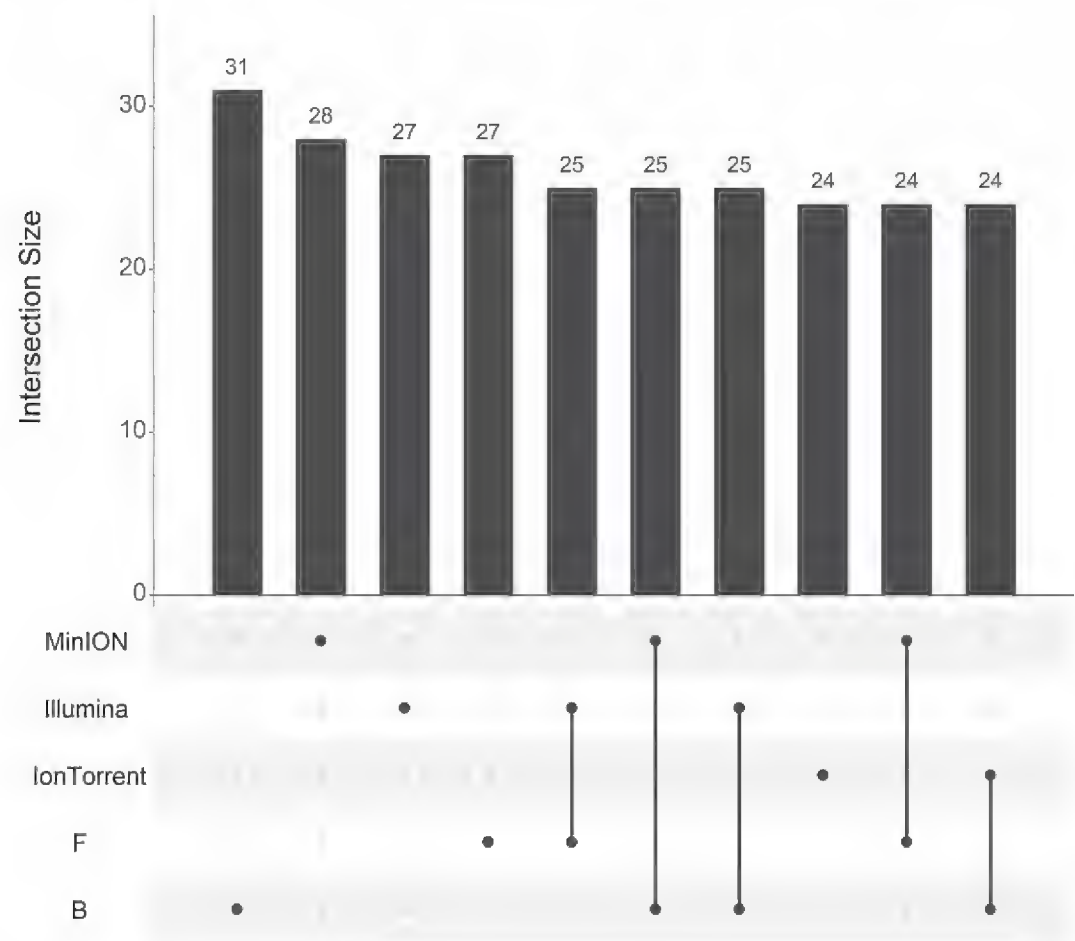


Figure 2. Upset plot showing the number of individuals detected at species level according to the three technologies (Illumina, MinION and IonTorrent), primer pairs (F=fwhF2/fwhR2n [254 bp] and B=BF3/BR2 [458 bp]) and technology-primer pair combinations tested.

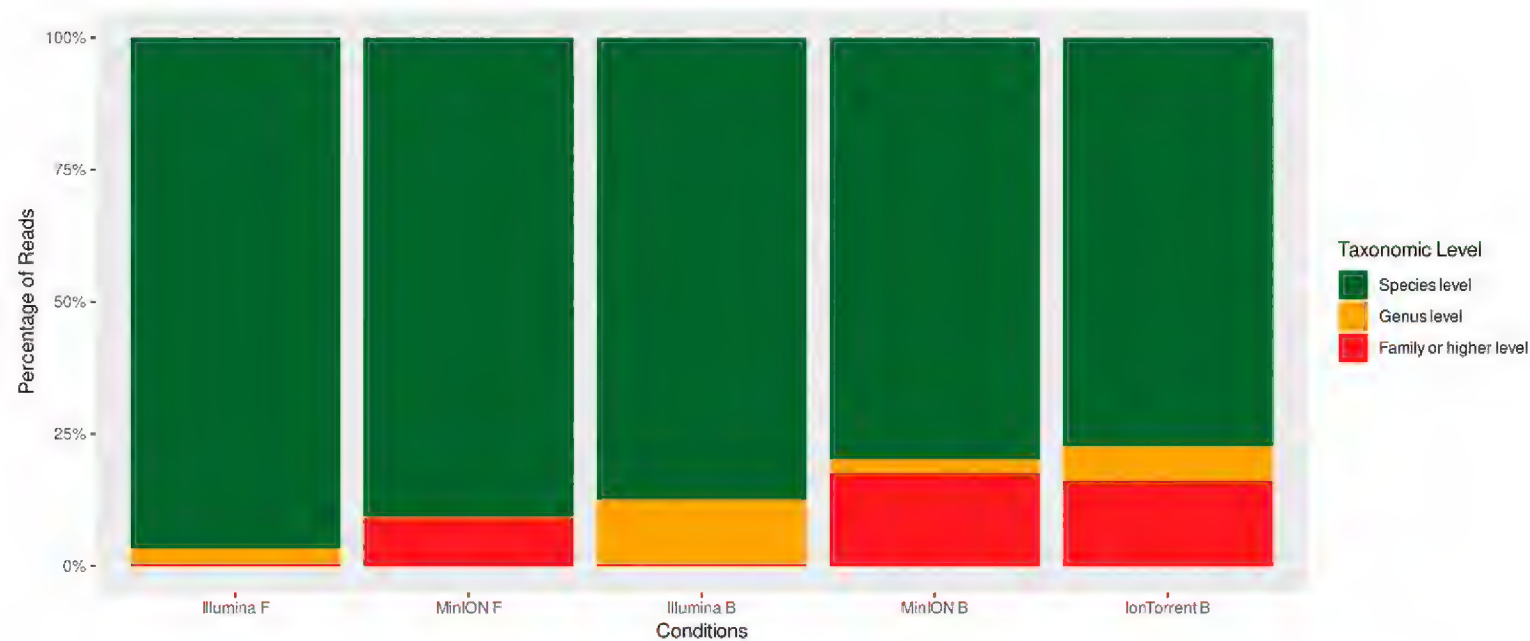


Figure 3. Proportion of reads assigned to each taxonomic level for each combination of sequencing technology and pair of primers (F: fwhF2/fwhR2n; B: BF3/BR2).

False positive detections (i.e. a species detected within a bulk that is not part of the bulk's initial composition) were observed regardless of the combination of primers and technology (Fig. 4). The precision (the ability to measures the proportion of correct detections) of the Illumina F and MinION F combinations are 0.625 and 0.667, respectively, while for the Illumina B, MinION B and IonTorrent B combinations, the precision is 0.806, 0.926 and 0.8 respectively. An average of 13.5 false positive OTUs were recorded for the primer pair fwH2/fwH2n, compared to an average of four false positive OTUs when using the primer pair BF3/BR2, the difference being significant here (p value = 0.00194). According to the technology used, but regardless of the primers, an average of ten, seven and six false positives were recorded for Illumina®, MinION and IonTorrent® technologies, respectively. There are no significant differences amongst the three sequencing technologies in terms of false positives.

Mock community analysis

In total, 33 out of 48 individuals (68.8%) were detected at the species level by at least one experimental condition (Fig. 5).

Bulks 1 and 2 were assembled to compare the detection rates of closely-related species under different sequencing and primer conditions. Illumina® detected seven species out of 16 (43.75%), MinION also detected seven out of 16 (43.75%) and IonTorrent® detected six species out of 16 (37.5%). No significant differences were observed amongst the different methods used (Kruskal-Wallis chi-squared = 2, df = 2, p value = 0.3679).

Metabarcoding of bulks 3 and 4 aimed at comparing the ability of different sequencing technologies to detect low abundance species in the traps. All sequencing technology/primer combinations allowed for the detection of minor species: *Phymatodes testaceus* with a presence of 3% in bulk 3 (relative amount of DNA in the mock community) and *Xylotrechus chinensis* with a percentage of 0.5% in bulk 4. However, some species (although not in minority in the bulks) were not detected in one or several tests (Fig. 5). In total, Illumina® was able to detect a higher number of individuals (11/12 individuals detected) compared to MinION (9/12) and IonTorrent® (9/12).

Regarding bulks mimicking the species composition in a field trap, MinION performed better to detect and identify specimens at the species level in Bulk 6

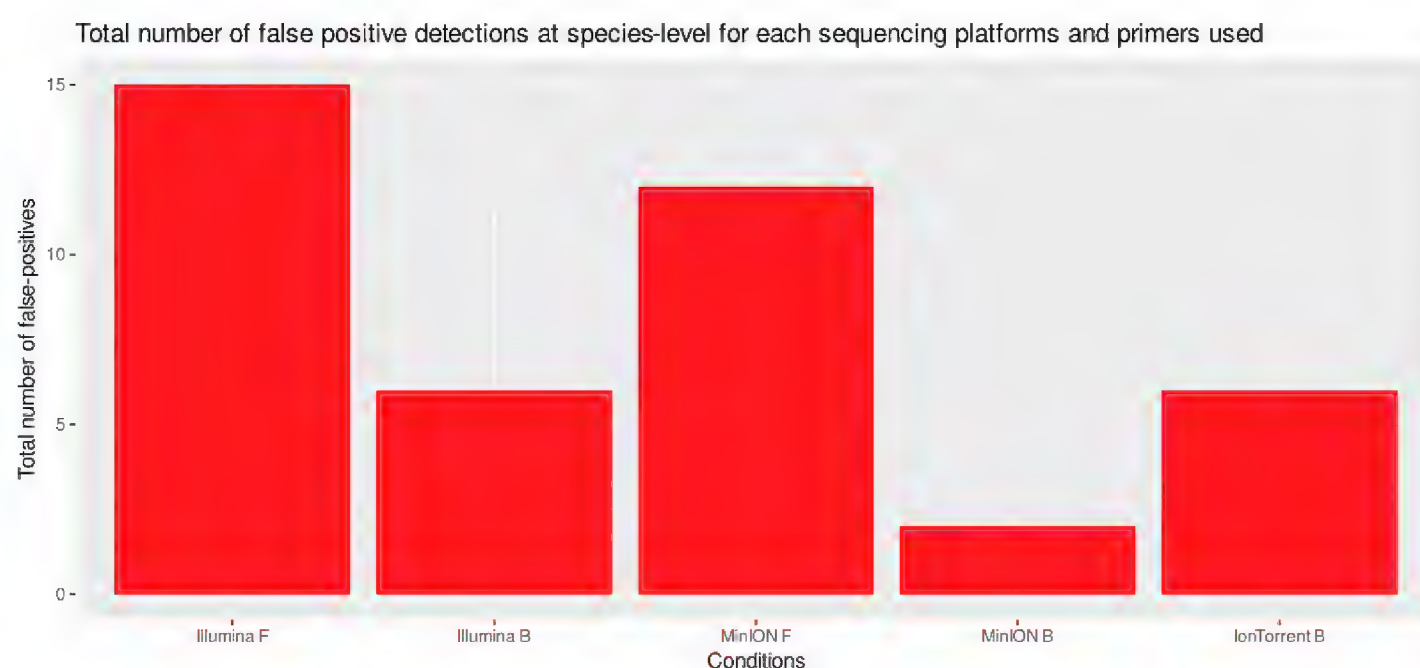


Figure 4. Number of false positive detections at species level for each sequencing platform and primers used (F=fwH2/fwH2n [254 bp] and B=BF3/BR2 [458 bp]).

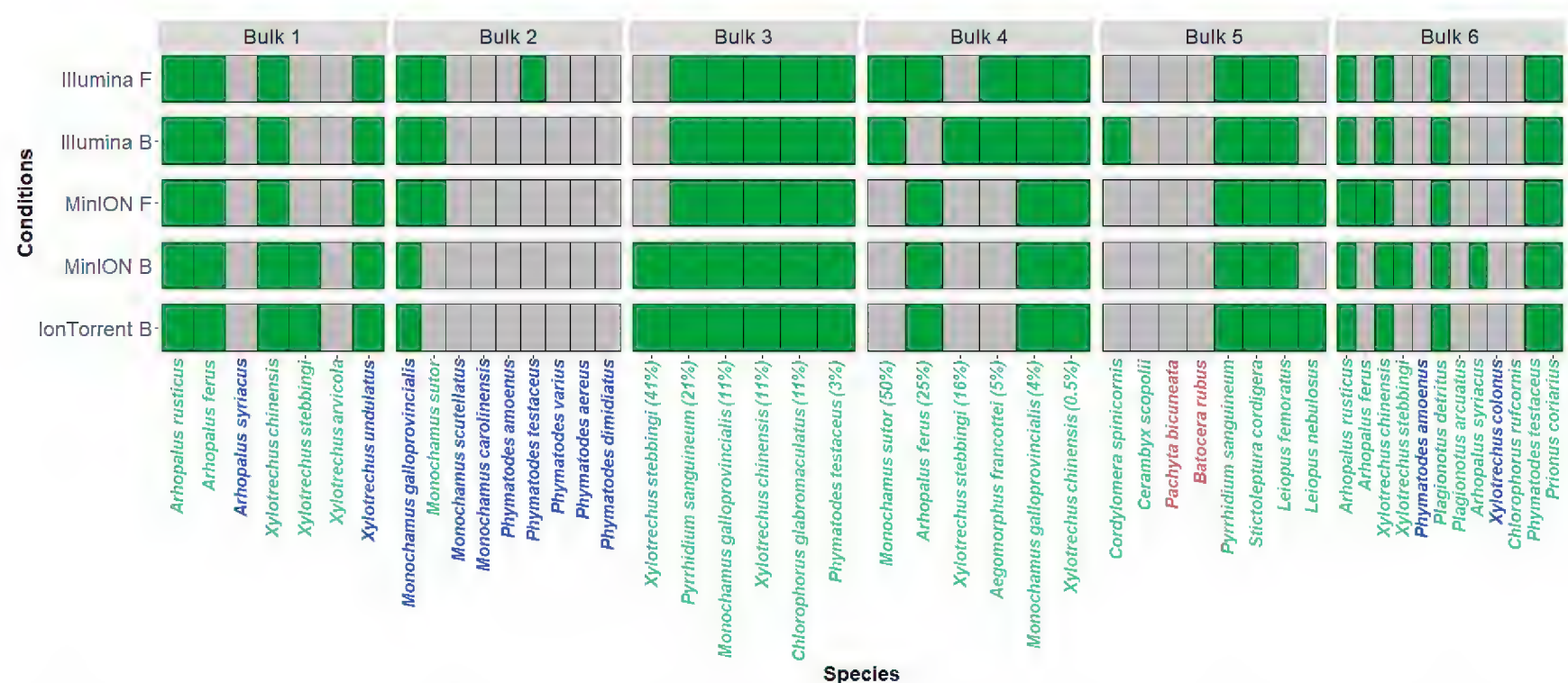


Figure 5. Heatmap comparing the identification of individuals present in bulk samples at the species level (green square) or the absence of detection at the species level (grey square) according to the sequencing technologies and primer pairs used (F=fwhF2/FwhR2n; B = BF3/BR2). Species names written in blue were collected using the wet method, those in green were collected using the dry method and those in dark red were hand-captured.

(detecting 8/12 species (66.7%)) compared to Illumina® and Ion Torrent® technologies (5/12 species (41.7%)), whereas the same number of species was detected for Bulk 5 (4/6 (66.7%)) regardless of the technology used. Nevertheless, in bulk 5, the non-native species, *Cordylomera spinicornis* was detected only by Illumina B. For bulk 6, the non-native species *Xylotrechus chinensis* was detected by all three technologies and *Xylotrechus stebbingi* by MinION B only.

Impact of capture and storage conditions on individual detection

Our results demonstrate significant differences in the mean number of detections between samples that were collected using the “dry” method (α -cypermethrin insecticide) and the “wet” method (water-diluted propylene glycol) (Wilcoxon rank-sum test, $W = 74.5$, $p \text{ value} = 0.0006342$) (Fig. 6A). Indeed, 75% (9/12) of specimens collected using the “wet” MPG trapping procedure were not detected by any of the sequencing technologies. Conversely, 88.2% (30/34) of those collected using the “dry” trapping procedure (based on α -cypermethrin insecticide) were detected at least once across all technologies.

Individuals captured using the “dry” method had higher DNA concentration (39 ng/ μ l on average (SD = 52.79)) than MPG trapped specimens (18.6 ng/ μ l on average (SD = 21.80)) (Wilcoxon rank-sum test, $W = 123.5$, $p \text{ value} = 0.04533$) (Fig. 6B). The average A260/280 ratio was 1.9 for the “dry” method and 2 for the MPG method (Fig. 6C). However, the average A260/230 ratio of specimens trapped with the “dry” method (0.8) was higher than that of specimens captured with MPG (0.5) (Wilcoxon rank-sum test, $W = 146$, $p \text{ value} = 0.1502$) (Fig. 6D).

Discussion

Rapid and precise detection of exotic insects is crucial to prevent the ecological and economic damage they can cause by invading new environments and disrupting local ecosystems.

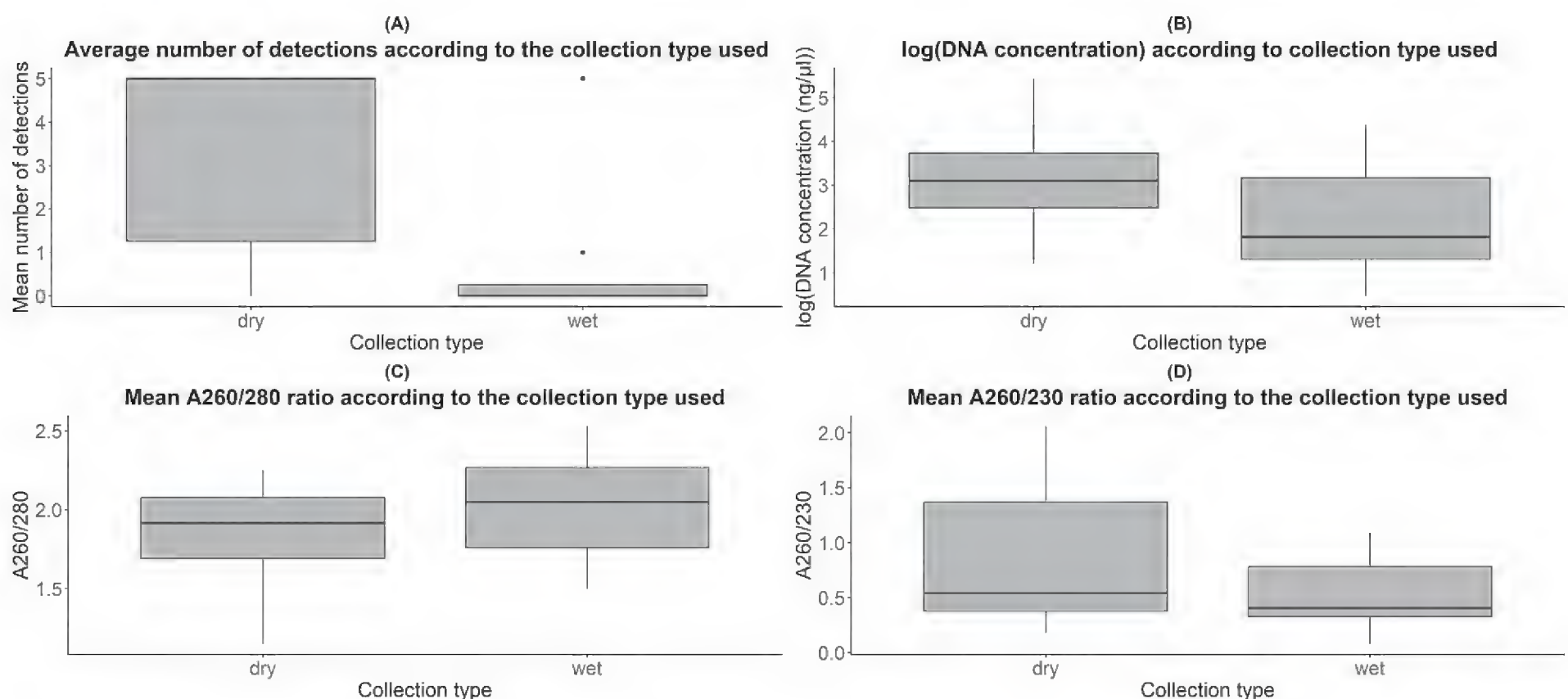


Figure 6. Boxplots representing (A) the average number of detections according to the type of preservation used, (B) the natural logarithm scale (base e) of the average DNA concentration according to the type of preservation used, (C) the A260/280 quality ratio according to the type of preservation used and (D) the A260/230 quality ratio according to the type of preservation used. The black dots represent the outlier values (values outside the whiskers). The bold line represents the average value, outlines of the boxes represent the first and third quartiles and the whiskers represent the range of the values outside the quartiles.

Benchmarking of sequencing technologies

A slightly higher number of individuals were detected and identified to species using MinION (28 specimens) compared to Illumina® (27 specimens) or IonTorrent® (24 specimens), although this difference is not significant. However, this result demonstrates that the sequencing error rates long attributed to the MinION did not impact detection rates, while allowing for the elimination of the long delays often required when sequencing is performed on other sequencing technologies (Piper et al. 2019). It must be considered that we worked on a single pair of primers (BF3/BR2) with the IonTorrent® technology, which may have reduced the number of identifications. More specifically, our results showed that the choice of primer pairs and the length of the amplicon generated led to contrasted results regarding taxonomic assignment. For example, only BF3/BR2 allowed the species-level identification of the invasive species *Xylotrechus stebbingi*. This difference may be due to the longer amplicon generated by this primer pair, which has more informative nucleotide sites to provide a reliable taxonomic assignment. By contrast, fwhF2/fwhR2n generated a significantly higher number of false positives than BF3/BR2 (Fig. 4). This may be because the amplicon generated by fwhF2/fwhR2n is smaller in size compared to BF3/BR2. As a result, any loss of genetic information is more likely to result in misidentification or false positives (Meusnier et al. 2008).

Regardless of the number of identified species, the Illumina® technology produced a higher percentage of reads allowing species-level identification compared to MinION or IonTorrent®. The detection of specimens at a higher taxonomic level (genus or family) can be explained by sequencing errors that produce reads with less than 98% identity to the reference database. These results confirm that Illumina has a lower sequencing error rate than Oxford Nanopore's MinION sequencer (Piper et al. 2019), although this did not impact the number of individuals identified to the species level.

The three technologies showed similar efficiency in detecting and identifying closely-related species. Moreover, the results show that all three sequencing technologies (regardless of the associated primer pairs) enabled the detection and identification of species whose DNA represented a very low percentage in the mock community (Fig. 5). This high resolution would allow for the detection of exotic species that are poorly represented in traps, which might otherwise go unnoticed. Thus, all three technologies appear suitable for detecting and identifying species present in low numbers in field traps, enabling effective monitoring.

Impacts of capture and storage conditions on DNA conservation

Both the conditions of capture (wet versus dry methods) and storage (i.e. time lag between collection and lab processing) have an impact on DNA concentration and quality and subsequently on the rate of species detection (Piper et al. 2019). Thus, the number of species detected is highly variable between bulks 1 and 2, which can be explained by the capture methods used: ‘dry,’ where individuals were captured without preservative fluid (as is the case for the majority of detected individuals comprising bulk 1) and ‘wet,’ where individuals were preserved in 50% MPG until trap retrieval (as is the case for the majority of undetected individuals comprising bulk 2). For instance, the species *Phymatodes testaceus* was always detected (10 out of 10 assays) when dry specimens were present, even in low concentrations (3% in Bulk 3). On the other hand, wet specimens of *P.testaceus* were rarely detected (one detection out of five assays). Individuals captured using the MPG method had lower DNA concentration and presented significantly lower detection rates compared to individuals captured using the “dry” method. Ballare et al. (2019) also found that insects collected in propylene glycol traps produced lower quality ddRADseq assemblages than specimens collected by net sampling and directly transferred into 100% ethanol (EtOH) or by passive trapping followed by 100% EtOH storage before pinning. In contrast to this, Ferro and Park (2013) found that propylene glycol is an effective DNA preservative for molecular marker-based studies on Coleoptera species. However, in their study, insects were first killed and preserved in 100% ethanol before being stored in glycol, while in our study insects were killed directly in propylene glycol. The use of 100% ethanol as the initial killing agent may lead to better initial preservation of specimens compared to direct exposure to propylene glycol.

False positives, negatives and unmatched OTUs

Despite the precautions taken, several false positives were detected in all tested conditions. The number of false positives was significantly higher with the primer pair fwhF2/fwhR2n, which generates a smaller size amplicon compared to BF3/BR2. Even though Illumina technology is known to have a lower sequencing error rate compared to MinION (Silvestre-Ryan and Holmes 2021), our study found 10 false positives generated by Illumina, while MinION produced seven false positives and IonTorrent produced six. The sensitivity of HTS technologies allows for the detection of very small amounts of DNA, thus detecting even the slightest cross-contamination between samples (Liu et al. 2020). These DNA contaminations may have occurred during sample collection in the field or in the laboratory through cross-contamination between samples from the same study.

The false negative detections for some individuals may primarily be explained by the highly heterogeneous DNA quality of the different sequenced individuals (Suppl. material 1: table S1). In fact, DNA quality can be impacted by numerous mainly abiotic factors (pH, UV radiation, temperature), degrading DNA quality in a matter of days or weeks (Strickler et al. 2015; Collins et al. 2018; Harrison et al. 2019). During field trapping using stationary traps, captured insects are sometimes exposed to such conditions (high temperatures in trap containers when exposed to the sun in summer, high humidity in the container during heavy rains etc...), which can greatly accelerate the speed of DNA degradation in captured individuals. Such degraded DNA is more difficult to amplify, thus generating false negatives, especially when attempting to detect insects in low abundance within a trap, such as an invasive species in the process of establishing (Preston et al. 2022). Another possible cause for the high number of false negatives is the bias induced by PCR, such as uneven amplification of the DNA of the different individuals present in one sample (Preston et al. 2022). To avoid potential bias arising from identification errors or missing species in the reference databases, we decided to work on a local and curated BLAST database. However, when target species are partially unknown, as is the case in field conditions, analyses must rely on public reference databases. Yet, out of the 35,000 known species of Cerambycidae to date, only 2,926 species (8.4%) are recorded in BOLD with a barcode fragment (as of 16 November 2023). Furthermore, databases can contain errors such as mis-assignment of a DNA sequence to a wrong species due to morphological identification errors. This was precisely the error encountered for the species *Monochamus sutor*, which was genetically identified as *Monochamus sartor* (Suppl. material 1: table S7) or the species *Leiopus nebulosus* which has been genetically identified as *Leiopus linnei* (Suppl. material 1: table S7) using our local BOLD database.

One also needs to pay attention to synonymy whereby a species appears in the database under multiple names. We encountered this problem in our analysis with *Arhopalus fesus* (Bulks 1, 4 and 6) which was detected, but under the name of *Arhopalus tristis* (Suppl. material 1: table S10). Finally, mitochondrial paralogues such as NUMTs (non-functional copies of mitochondrial genes transported into the nuclear genome) present in databases can also bias results, making it impossible to identify specimens correctly at the species level (Bensasson et al. 2001). NUMTs are numerous in many organisms, including some cerambycids such as *Monochamus galloprovincialis* (Koutroumpa et al. 2009; Haran et al. 2015).

The differences in identification or non-detection between morphologically similar species belonging to the same genus, as observed, for example, with *Monochamus* spp., *Phymatodes* spp. or *Arhopalus* spp. (Fig. 5), can be explained in part by the reasons mentioned above.

Biases

Based on the results obtained, it appears that the main biases observed in metabarcoding analyses of trap contents stem from the degradation of DNA from individuals, which generates false negatives. We recommend favouring a “dry” rather than a “wet” trapping method, especially the MPG method and to plan for the collection, transportation and processing of captured individuals as soon as possible after capture. This includes checking the traps as frequently as possible (at least once a week), thus avoiding excessively long exposure of the individuals

to unfavourable environmental conditions. Once individuals are brought back to the laboratory and if DNA cannot be extracted straight away, it is important to limit any further degradation by keeping samples at -20 °C and in 95% ethanol. On the other hand, DNA extractions should be stored in the preservation buffer provided with the extraction kits or in molecular-grade water and kept at -20 °C (Preston et al. 2022). We also recommend limiting the use of primer pairs that generate short amplicons, which can favour the amplification of non-target taxa, NUMTs and lead to identification errors. The quality and completeness of the databases are also very important bias factors. To limit this bias, Egeter et al. (2022) recommended restricting the database used to targeted species in order to minimise the risk of false positives due to contamination. Limited taxonomic and geographical coverage of sequence databases is a huge limitation in metabarcoding studies. For example, Dopheide et al. (2019) found no representative sequence in the GenBank database for more than 900 invertebrate OTUs in their study when analysing the community of soil arthropods from a native forest in Ireland. Additionally, species identification errors and cases of synonymy lead to false negatives or cases of multiple affiliations.

Conclusion

By comparing the accuracy and detection capacity of three metabarcoding strategies, this study contributes to improving our toolkit for monitoring non-native insect invasion. All three sequencing technologies performed equally well and showed similar results for detecting and identifying exotic Cerambycid species collected in field traps. However, MinION stands out as a portable, easy-to-use, and cost-effective sequencer, with the potential to become an essential tool for biodiversity monitoring projects. Using MinION reduces the time spent on laboratory handling compared to Illumina and eliminates the need to outsource sample sequencing. This saves considerable time when it comes to detecting invasive species. The MinION technology is accurate enough to detect non-native species even when present at low abundances in field traps and allows for accurate identifications as long as there is a sufficiently complete high-quality reference database to avoid identification errors or false positives/negatives. It is also crucial to pay close attention to issues of contamination and specimen preservation during and after individual capture in order to work with the least degraded DNA possible.

Acknowledgements

We would like to thank all colleagues who participated in the taxa sampling (see Roques et al. (2023)). We are thankful to Lucas Sire for insightful discussion on primer choice. The authors thank the GeT-Biopuces platform of INSA Toulouse for the IonTorrent sequencing study. We would like to thank the Mutualized Platform for Environmental Genomics of the Institut de Recherche sur la Biologie de l'Insecte, UMR 7261 CNRS, for access to its facilities and logistic support.

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Funding

This work was supported by the PORTRAP project “Test de l’efficacité de pièges génériques multi-composés pour la détection précoce d’insectes exotiques xylophages dans les sites potentiels d’entrée sur le territoire national” and HOMED project (HOListic Management of Emerging Forest Pests and Diseases) which received funding from the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No. 771271 (<https://homed-project.eu/>). We are grateful to the genotoul bioinformatics platform Toulouse Midi-Pyrenees for providing help and computing storage resources. Loïs Veillat was supported by a PhD studentship from HOMED project and doctoral school SSBCV at the University of Orléans.

Author contributions

Loïs Veillat, Géraldine Roux, Carlos Lopez-Vaamonde and Stéphane Boyer conceived the study. Alain Roques collected field samples. Stéphane Boyer, Marina Querejeta, Emmanuelle Magnoux and Loïs Veillat conducted the laboratory sample processing. Loïs Veillat analysed the data and wrote the first draft. All authors contributed to the preparation of the manuscript. Both senior authors, Géraldine Roux and Carlos Lopez-Vaamonde, contributed equally to this study.

Author ORCIDs

Loïs Veillat  <https://orcid.org/0009-0004-2149-1336>

Stéphane Boyer  <https://orcid.org/0000-0002-0750-4864>

Marina Querejeta  <https://orcid.org/0000-0003-1803-5239>

Emmanuelle Magnoux  <https://orcid.org/0000-0003-0990-5511>

Alain Roques  <https://orcid.org/0000-0002-3734-3918>

Carlos Lopez-Vaamonde  <https://orcid.org/0000-0003-2278-2368>

Géraldine Roux  <https://orcid.org/0000-0002-1116-2799>

Data availability

Barcode data for the 33 species used in the mock community experiment are available from BOLD in the dataset DS-MINION (dx.doi.org/10.5883/DS-MINION). Raw sequence data for this project and analytical script and files are available on figshare (https://figshare.com/projects/DNA_metabarcoding_an_efficient_way_to_detect_non-native_cerambycid_beetles_in_trapping_collections_/171432).

References

- Abeynayake SW, Fiorito S, Dinsdale A, Whattam M, Crowe B, Sparks K, Campbell PR, Gambley C (2021) A Rapid and Cost-Effective Identification of Invertebrate Pests at the Borders Using MINION Sequencing of DNA Barcodes. *Genes* 12(8): 1138. <https://doi.org/10.3390/genes12081138>
- Alberdi A, Aizpurua O, Gilbert MTP, Bohmann K (2018) Scrutinizing key steps for reliable metabarcoding of environmental samples. Mahon A (Ed.) *Methods in Ecology and Evolution* 9: 134–147. <https://doi.org/https://doi.org/10.1111/2041-210X.12849>
- Andújar C, Arribas P, Gray C, Bruce C, Woodward G, Yu DW, Vogler AP (2018) Metabarcoding of freshwater invertebrates to detect the effects of a pesticide spill. *Molecular Ecology* 27(1): 146–166. <https://doi.org/10.1111/mec.14410>
- Aukema JE, McCullough DG, Von Holle B, Liebhold AM, Britton K, Frankel SJ (2010) Historical Accumulation of Nonindigenous Forest Pests in the Continental United States. *Bioscience* 60(11): 886–897. <https://doi.org/10.1525/bio.2010.60.11.5>

- Ballare KM, Pope NS, Castilla AR, Cusser S, Metz RP, Jha S (2019) Utilizing field collected insects for next generation sequencing: Effects of sampling, storage, and DNA extraction methods. *Ecology and Evolution* 9(24): 13690–13705. <https://doi.org/10.1002/ece3.5756>
- Batovska J, Piper AM, Valenzuela I, Cunningham JP, Blacket MJ (2021) Developing a non-destructive metabarcoding protocol for detection of pest insects in bulk trap catches. *Scientific Reports* 11(1): 7946. <https://doi.org/10.1038/s41598-021-85855-6>
- Bensasson D, Zhang D-X, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: Evolution's misplaced witnesses. *Trends in Ecology & Evolution* 16(6): 314–321. [https://doi.org/10.1016/S0169-5347\(01\)02151-6](https://doi.org/10.1016/S0169-5347(01)02151-6)
- Blackburn TM, Pyšek P, Bacher S, Carlton JT, Duncan RP, Jarošík V, Wilson JR, Richardson DM (2011) A proposed unified framework for biological invasions. *Trends in Ecology & Evolution* 26(7): 333–339. <https://doi.org/10.1016/j.tree.2011.03.023>
- Bradshaw CJA, Leroy B, Bellard C, Roiz D, Albert C, Fournier A, Barbet-Massin M, Salles J-M, Simard F, Courchamp F (2016) Massive yet grossly underestimated global costs of invasive insects. *Nature Communications* 7(1): 12986. <https://doi.org/10.1038/ncomms12986>
- Braukmann TWA, Ivanova NV, Prosser SWJ, Elbrecht V, Steinke D, Ratnasingham S, De Waard JR, Sones JE, Zakharov EV, Hebert PDN (2019) Metabarcoding a diverse arthropod mock community. *Molecular Ecology Resources* 19(3): 711–727. <https://doi.org/10.1111/1755-0998.13008>
- Brown SDJ, Collins RA, Boyer S, Lefort M, Malumbres-Olarte J, Vink CJ, Cruickshank RH (2012) S PIDER: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources* 12(3): 562–565. <https://doi.org/10.1111/j.1755-0998.2011.03108.x>
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: Architecture and applications. *BMC Bioinformatics* 10(1): 421. <https://doi.org/10.1186/1471-2105-10-421>
- Chain FJJ, Brown EA, MacIsaac HJ, Cristescu ME (2016) Metabarcoding reveals strong spatial structure and temporal turnover of zooplankton communities among marine and freshwater ports. Cowie R (Ed.) *Diversity and Distributions* 22: 493–504. <https://doi.org/10.1111/ddi.12427>
- Chown SL, Hodgins KA, Griffin PC, Oakeshott JG, Byrne M, Hoffmann AA (2015) Biological invasions, climate change and genomics. *Evolutionary Applications* 8(1): 23–46. <https://doi.org/10.1111/eva.12234>
- Chua PYS, Bourlat SJ, Ferguson C, Korlevic P, Zhao L, Ekrem T, Meier R, Lawniczak MKN (2023) Future of DNA-based insect monitoring. *Trends in Genetics* 39(7): 531–544. <https://doi.org/10.1016/j.tig.2023.02.012>
- Collins RA, Wangenstein OS, O’Gorman EJ, Mariani S, Sims DW, Genner MJ (2018) Persistence of environmental DNA in marine systems. *Communications Biology* 1(1): 185. <https://doi.org/10.1038/s42003-018-0192-6>
- deWaard JR, Levesque-Beaudin V, deWaard SL, Ivanova NV, McKeown JTA, Miskie R, Naik S, Perez KHJ, Ratnasingham S, Sobel CN, Sones JE, Steinke C, Telfer AC, Young AD, Young MR, Zakharov EV, Hebert PDN (2019) Expedited assessment of terrestrial arthropod diversity by coupling Malaise traps with DNA barcoding 1. *Genome* 62(3): 85–95. <https://doi.org/10.1139/gen-2018-0093>
- Dopheide A, Tooman LK, Grosser S, Agabiti B, Rhode B, Xie D, Stevens MI, Nelson N, Buckley TR, Drummond AJ, Newcomb RD (2019) Estimating the biodiversity of terrestrial invertebrates on a forested island using DNA barcodes and metabarcoding data. *Ecological Applications* 29: 1–14. <https://doi.org/10.1002/eap.1877>
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME Improves Sensitivity and Speed of Chimera Detection. *Bioinformatics (Oxford, England)* 27(16): 2194–2200. <https://doi.org/10.1093/bioinformatics/btr381>

- Egeter B, Peixoto S, Brito JC, Jarman S, Puppo P, Velo-Antón G (2018) Challenges for assessing vertebrate diversity in turbid Saharan water-bodies using environmental DNA. *Genome* 61(11): 807–814. <https://doi.org/10.1139/gen-2018-0071>
- Egeter B, Veríssimo J, Lopes-Lima M, Chaves C, Pinto J, Riccardi N, Beja P, Fonseca NA (2022) Speeding up the detection of invasive bivalve species using environmental DNA: A Nanopore and Illumina sequencing comparison. *Molecular Ecology Resources* 22(6): 2232–2247. <https://doi.org/10.1111/1755-0998.13610>
- Elbrecht V, Leese F (2017) Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment. *Frontiers in Environmental Science* 5. <https://www.frontiersin.org/articles/10.3389/fenvs.2017.00011>
- Elbrecht V, Vamos EE, Meissner K, Aroviita J, Leese F (2017) Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. Yu D (Ed.) *Methods in Ecology and Evolution* 8: 1265–1275. <https://doi.org/10.1111/2041-210X.12789>
- Elbrecht V, Braukmann TWA, Ivanova NV, Prosser SWJ, Hajibabaei M, Wright M, Zakharov EV, Hebert PDN, Steinke D (2019) Validation of COI metabarcoding primers for terrestrial arthropods. *PeerJ* 7: e7745. <https://doi.org/10.7717/peerj.7745>
- Escudié F, Auer L, Bernard M, Mariadassou M, Cauquil L, Vidal K, Maman S, Hernandez-Raquet G, Combes S, Pascal G (2018) FROGS: Find, Rapidly, OTUs with Galaxy Solution. Berger B (Ed.) *Bioinformatics* 34: 1287–1294. <https://doi.org/10.1093/bioinformatics/btx791>
- Ferro ML, Park J-S (2013) Effect of Propylene Glycol Concentration on Mid-Term DNA Preservation of Coleoptera. In: *The Coleopterists Bulletin*, 581–586. <https://doi.org/10.1649/0010-065X-67.4.581>
- Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics (Oxford, England)* 28(23): 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>
- Gibson J, Shokralla S, Porter TM, King I, van Konynenburg S, Janzen DH, Hallwachs W, Hajibabaei M (2014) Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasytematics. *Proceedings of the National Academy of Sciences of the United States of America* 111(22): 8007–8012. <https://doi.org/10.1073/pnas.1406468111>
- Giovani B, Blümel S, Lopian R, Teulon D, Bloem S, Galeano Martínez C, Beltrán Montoya C, Urias Morales CR, Dharmapuri S, Timote V, Horn N, Chouibani M, Mezui M’Ella JG, Herrera V, Castinel A, Goletsos C, Moeller C, Naumann I, Stancanelli G, Bronzwaer S, Tramontini S, MacDonald P, Matheson L, Anthoine G, De Jonghe K, Schenk M, Steinmüller S, Rodriguez E, Cruz ML, Luck J, Fraser G, Brunel S, Montuori M, Fedchock C, Steel E, Pennington HG, Day R, Rossi JP, Xia J (2020) Science diplomacy for plant health. *Nature Plants* 6(8): 902–905. <https://doi.org/10.1038/s41477-020-0744-x>
- Haran J, Koutroumpa F, Magnoux E, Roques A, Roux G (2015) Ghost mtDNA haplotypes generated by fortuitous NUMTs can deeply disturb infra-specific genetic diversity and phylogeographic pattern. *Journal of Zoological Systematics and Evolutionary Research* 53(2): 109–115. <https://doi.org/10.1111/jzs.12095>
- Harrison JB, Sunday JM, Rogers SM (2019) Predicting the fate of eDNA in the environment and implications for studying biodiversity. *Proceedings. Biological Sciences* 286(1915): 20191409. <https://doi.org/10.1098/rspb.2019.1409>
- Hebert PDN, Ratnasingham S, deWaard JR (2003) Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings. Biological Sciences* 270(suppl_1): S96–S99. <https://doi.org/10.1098/rsbl.2003.0025>
- Henrie JR, Thomson BM, Yungfleisch AA, Kerr M, Leavitt SD (2022) Characterizing Crustose Lichen Communities—DNA Metabarcoding Reveals More than Meets the Eye. *Diversity* 14(9): 766. <https://doi.org/10.3390/d14090766>

- Ho JKI, Puniamoorthy J, Srivathsan A, Meier R (2020) MinION sequencing of seafood in Singapore reveals creatively labelled flatfishes, confused roe, pig DNA in squid balls, and phantom crustaceans. *Food Control* 112: 107144. <https://doi.org/10.1016/j.foodcont.2020.107144>
- Hodgetts J, Ostojá-Starzewski JC, Prior T, Lawson R, Hall J, Boonham N (2016) DNA barcoding for biosecurity: case studies from the UK plant protection program. Wilson J-J (Ed.) *Genome* 59: 1033–1048. <https://doi.org/10.1139/gen-2016-0010>
- Javal M, Terblanche J, Conlong D, Delahaye N, Grobbelaar E, Benoit L, Lopez-Vaamonde C, Haran J (2021) DNA barcoding for bio-surveillance of emerging pests and species identification in Afrotropical Prioninae (Coleoptera, Cerambycidae). *Biodiversity Data Journal* 9: e64499. <https://doi.org/10.3897/BDJ.9.e64499>
- Kelnarova I, Jendek E, Grebennikov VV, Bocak L (2019) First molecular phylogeny of *Agrilus* (Coleoptera: Buprestidae), the largest genus on Earth, with DNA barcode database for forestry pest diagnostics. *Bulletin of Entomological Research* 109(2): 200–211. <https://doi.org/10.1017/S0007485318000330>
- Kenis M, Auger-Rozenberg M-A, Roques A, Timms L, Péré C, Cock MJW, Settele J, Augustin S, Lopez-Vaamonde C (2009) Ecological effects of invasive alien insects. *Biological Invasions* 11(1): 21–45. <https://doi.org/10.1007/s10530-008-9318-y>
- Kitson JJN, Hahn C, Sands RJ, Straw NA, Evans DM, Lunt DH (2019) Detecting host-parasitoid interactions in an invasive Lepidopteran using nested tagging DNA metabarcoding. *Molecular Ecology* 28(2): 471–483. <https://doi.org/10.1111/mec.14518>
- Koutroumpa FA, Lieutier F, Roux-Morabito G (2009) Incorporation of mitochondrial fragments in the nuclear genome (*Numts*) of the longhorned beetle *Monochamus galloprovincialis* (Coleoptera, Cerambycidae). *Journal of Zoological Systematics and Evolutionary Research* 47(2): 141–148. <https://doi.org/10.1111/j.1439-0469.2008.00492.x>
- Krehenwinkel H, Pomerantz A, Prost S (2019) Genetic Biomonitoring and Biodiversity Assessment Using Portable Sequencing Technologies: Current Uses and Future Directions. *Genes* 10(11): 858. <https://doi.org/10.3390/genes10110858>
- Liu M, Clarke LJ, Baker SC, Jordan GJ, BurrIDGE CP (2020) A practical guide to DNA metabarcoding for entomological ecologists. *Ecological Entomology* 45(3): 373–385. <https://doi.org/10.1111/een.12831>
- Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M (2014) Swarm: Robust and fast clustering method for amplicon-based studies. *PeerJ* 2: e593. <https://doi.org/10.7717/peerj.593>
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* 17(1): 10. <https://doi.org/10.14806/ej.17.1.200>
- Mas H, Santoiemma G, Lencina JL, Gallego D, Pérez-Laorga E, Ruzzier E, Rassati D (2023) Investigating beetle communities in and around entry points can improve surveillance at national and international scale. *NeoBiota* 85: 145–165. <https://doi.org/10.3897/neobiota.85.103904>
- Mazza G, Tricarico E, Genovesi P, Gherardi F (2014) Biological invaders are threats to human health: An overview. *Ethology Ecology and Evolution* 26(2–3): 112–129. <https://doi.org/10.1080/03949370.2013.863225>
- Meusnier I, Singer GAC, Landry J-F, Hickey DA, Hebert PDN, Hajibabaei M (2008) A Universal DNA Mini-Barcode for Biodiversity Analysis. *BMC Genomics* 9(1): 214. <https://doi.org/10.1186/1471-2164-9-214>
- Piper AM, Batovska J, Cogan NOI, Weiss J, Cunningham JP, Rodoni BC, Blacket MJ (2019) Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance. *GigaScience* 8(8): giz092. <https://doi.org/10.1093/gigascience/giz092>
- Preston M, Fritzsche M, Woodcock P (2022). Understanding and mitigating errors and biases in metabarcoding: an introduction for non-specialists. JNCC Report No. 699, JNCC, Peterborough.
- Ratnasingham S, Hebert PDN (2007) BARCODING: bold: The Barcode of Life Data System (<http://www.barcodinglife.org>): BARCODING. *Molecular Ecology Notes* 7: 355–364. <https://doi.org/10.1111/j.1471-8286.2007.01678.x>

- Richardson DM, Pysek P, Rejmanek M, Barbour MG, Panetta FD, West CJ (2000) Naturalization and invasion of alien plants: Concepts and definitions. *Diversity. Diversity & Distributions* 6(2): 93–107. <https://doi.org/10.1046/j.1472-4642.2000.00083.x>
- Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: A Versatile Open Source Tool for Metagenomics. *PeerJ* 4(Oct): e2584. <https://doi.org/10.7717/peerj.2584>
- Roques A, Auger-Rozenberg M-A, Blackburn TM, Garnas J, Pyšek P, Rabitsch W, Richardson DM, Wingfield MJ, Liebhold AM, Duncan RP (2016) Temporal and interspecific variation in rates of spread for insect species invading Europe during the last 200 years. *Biological Invasions* 18(4): 907–920. <https://doi.org/10.1007/s10530-016-1080-y>
- Roques A, Ren L, Rassati D, Shi J, Akulov E, Audsley N, Auger-Rozenberg M-A, Avtzis D, Battisti A, Bellanger R, Bernard A, Bernadinelli I, Branco M, Cavaletto G, Cocquempot C, Contarini M, Courtial B, Courtin C, Denux O, Dvořák M, Fan J, Feddern N, Francese J, Franzen EKL, Garcia A, Georgiev G, Georgieva M, Giarruzzo F, Gossner M, Gross L, Guarneri D, Hoch G, Hölling D, Jonsell M, Kirichenko N, Loomans A, Luo Y, McCullough D, Maddox C, Magnoux E, Marchioro M, Martinek P, Mas H, Mériguet B, Pan Y, Phélut R, Pineau P, Ray AM, Roques O, Ruiz M-C, Sarto i Monteys V, Speranza S, Sun J, Sweeney JD, Touroult J, Valladares L, Veillat L, Yuan Y, Zalucki MP, Zou Y, Žunič-Kosi A, Hanks LM, Millar JG (2023) Worldwide tests of generic attractants, a promising tool for early detection of non-native cerambycid species. *NeoBiota* 84: 169–209. <https://doi.org/10.3897/neobiota.84.91096>
- Rossa R, Goczał J (2021) Global diversity and distribution of longhorn beetles (Coleoptera: Cerambycidae). *The European Zoological Journal* 88(1): 289–302. <https://doi.org/10.1080/24750263.2021.1883129>
- Sahlin K, Medvedev P (2020) De Novo Clustering of Long-Read Transcriptome Data Using a Greedy, Quality Value-Based Algorithm. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 27: 472–484. <https://doi.org/10.1089/cmb.2019.0299>
- Seebens H, Blackburn TM, Dyer EE, Genovesi P, Hulme PE, Jeschke JM, Pagad S, Pyšek P, van Kleunen M, Winter M, Ansong M, Arianoutsou M, Bacher S, Blasius B, Brockerhoff EG, Brundu G, Capinha C, Causton CE, Celesti-Grapow L, Dawson W, Dullinger S, Economo EP, Fuentes N, Guénard B, Jäger H, Kartesz J, Kenis M, Kühn I, Lenzner B, Liebhold AM, Mosena A, Moser D, Nentwig W, Nishino M, Pearman D, Pergl J, Rabitsch W, Rojas-Sandoval J, Roques A, Rorke S, Rossinelli S, Roy HE, Scalera R, Schindler S, Štajerová K, Tokarska-Guzik B, Walker K, Ward DF, Yamanaka T, Essl F (2018) Global rise in emerging alien species results from increased accessibility of new source pools. *Proceedings of the National Academy of Sciences* 115(10): E2264–E2273. <https://doi.org/10.1073/pnas.1719429115>
- Seebens H, Bacher S, Blackburn TM, Capinha C, Dawson W, Dullinger S, Genovesi P, Hulme PE, Kleunen M, Kühn I, Jeschke JM, Lenzner B, Liebhold AM, Pattison Z, Pergl J, Pyšek P, Winter M, Essl F (2021) Projecting the continental accumulation of alien species through to 2050. *Global Change Biology* 27(5): 970–982. <https://doi.org/10.1111/gcb.15333>
- Silvestre-Ryan J, Holmes I (2021) Pair consensus decoding improves accuracy of neural network basecallers for nanopore sequencing. *Genome Biology* 22(1): 38. <https://doi.org/10.1186/s13059-020-02255-1>
- Srivathsan A, Lee L, Katoh K, Hartop E, Kutty SN, Wong J, Yeo D, Meier R (2021) ONTbarcoder and MinION barcodes aid biodiversity discovery and identification by everyone, for everyone. *BMC Biology* 19(1): 217. <https://doi.org/10.1186/s12915-021-01141-x>
- Srivathsan A, Feng V, Suárez D, Emerson B, Meier R (2024) ONTbarcoder 2.0: Rapid Species Discovery and Identification with Real-time Barcoding Facilitated by Oxford Nanopore R10. 4. *Cladistics* 40(2): 192–203. <https://doi.org/10.1111/cla.12566>

- Strickler KM, Fremier AK, Goldberg CS (2015) Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. *Biological Conservation* 183: 85–92. <https://doi.org/10.1016/j.biocon.2014.11.038>
- Vamos E, Elbrecht V, Leese F (2017) Short COI markers for freshwater macroinvertebrate metabarcoding. *Metabarcoding and Metagenomics* 1: e14625. <https://doi.org/10.3897/mbmg.1.14625>
- Vaser R, Sović I, Nagarajan N, Šikić M (2017) Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research* 27(5): 737–746. <https://doi.org/10.1101/gr.214270.116>
- Wang Q (2017) *Cerambycidae of the World: Biology and Pest Management*. CRC Press, 643 pp.
- Wu Y, Trepanowski NE, Molongoski JJ, Reagel PF, Lingafelter SW, Nadel H, Myers SW, Ray AM (2017) Identification of wood-boring beetles (Cerambycidae and Buprestidae) intercepted in trade-associated solid wood packaging material using DNA barcoding and morphology. *Scientific Reports* 7(1): 40316. <https://doi.org/10.1038/srep40316>
- Zenker MM, Rougerie R, Teston JA, Laguerre M, Pie MR, Freitas AVL (2016) Fast Census of Moth Diversity in the Neotropics: A Comparison of Field-Assigned Morphospecies and DNA Barcoding in Tiger Moths. *Cimmaruta R (Ed.) PLoS ONE* 11: e0148423. <https://doi.org/10.1371/journal.pone.0148423>

Supplementary material 1

Additional data (OTU tables, samples metadata; summary tables; ...)

Authors: Loïs Veillat, Stéphane Boyer, Marina Querejeta, Emmanuelle Magnoux, Alain Roques, Carlos Lopez-Vaamonde, Geraldine Roux

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (<http://opendata-commons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: <https://doi.org/10.3897/neobiota.96.130195.suppl1>